Cluster randomised trials in maternal and child health: implications for power and sample size

Richard Reading, Ian Harvey, Mary Mclean, on behalf of the CAB and Family Health Study Team

Abstract

Background—Interventions based in the community can be evaluated by randomising clusters, such as general practices, rather than individuals, as in conventional randomised trials. This increases the sample size needed because of intracluster correlation.

Aims—To estimate sample size requirements for cluster randomised trials of interventions based in general practice directed at common health problems affecting mothers and infants.

Methods—Data were collected from a pilot trial of the effect of Citizen's Advice Bureau services involving six general practices. Outcome measures included the Edinburgh postnatal depression score, the Warwick child health and morbidity profile, number of visits to the general practitioner, and two questionnaires delivered at the beginning and end of the study. Intracluster correlation coefficients and inflation factors (the ratio of the sample size required for a cluster randomised trial to that required for an individually randomised trial) were calculated.

Results—Intracluster correlation coefficients ranged from 0 (sleeping problems, accidental injury, hospitalisation) to 0.09 (maternal smoking), with most being < 0.04 (for example, maternal depression, breast feeding, general health, minor illness, behavioural problems, and visits to the general practitioner). Assuming 50 cases/practice, cluster randomised trials require sample sizes up to 3 times greater than individually randomised trials for most health outcomes measured.

Conclusions—These data enable sample sizes to be estimated for cluster randomised trials into a range of maternal and child health outcomes. Using such a design, approximately 40 practices would be sufficient to evaluate the effect of an intervention on maternal depression, sleeping, and behavioural problems, and non-routine visits to the general practitioner.

(Arch Dis Child 2000;82:79–83)

Keywords: cluster randomisation; primary care; sample size; intracluster correlation; maternal and child health

includes evaluations of many other modalities of treatment, including social and psychological interventions.^{1 2} However, where the intervention is directed at a whole community, randomisation of individuals to treatment or control might be impossible or impractical. In these circumstances, randomisation of communities (or clusters) is an alternative that retains the methodological strength of randomisation.^{3 4} Communities in this context can be clinics, general practices, schools, neighbourhoods, urban districts, or even whole towns and cities. In this paper we concentrate on randomisation of general practices.

Cluster randomised trials are an important method of evaluating community based interventions.⁵ Paediatricians need to be aware of their strengths and weaknesses because they can be widely applied to evaluative research in maternal and child health. The practical advantages are that whole practice populations can be studied; the organisation of the trial might be simpler; primary care workers might find this method of randomisation less intrusive; and the practical problems of offering an intervention to some, but not others, within a practice are overcome.

The theoretical advantages are that contamination of the effects of the intervention from subjects to controls is minimised,⁶ an important consideration for trials of health promotion interventions. In addition, this method is the only way of evaluating an intervention that is designed to influence the way a primary health care worker or team operates. Finally, it might be possible to evaluate interventions at the cluster level that could not ethically be introduced on an individually randomised basis.⁷ For all these reasons, cluster randomised trials are likely to be used much more frequently in the future.

However, there are disadvantages to this type of study design.^{3 8} For example, it might be difficult, or even impossible, to blind subjects and researchers. The decision of whether to recruit individuals within practices or to enrol whole practices en bloc throws up many ethical and practical dilemmas.⁷ Here, we consider the specific issue of the potential loss of statistical power caused by randomising groups rather than individuals, and how this affects sample size considerations. This is not simply an obscure statistical problem, but a fundamental aspect of the study design of such trials.⁹

We have conducted a pilot study for a cluster randomised trial of the effects on child and maternal health of providing a Citizen's Advice Bureau service within primary care. Citizen's

School of Health Policy and Practice, University of East Anglia, Norwich NR4 7TJ, UK
R Reading
I Harvey
M Mclean

Correspondence to:
Dr Reading
email r.reading@uea.ac.uk

Accepted 19 July 1999

In general, randomised controlled trials are thought to be the gold standard for evaluation of interventions to improve health. The design evolved from trials of drug treatment and now

Advice Bureaus in the UK are a network of organisations in which trained volunteers provide advice to clients on welfare, legal, employment, housing, and consumer rights; on financial and debt management; on resolution of disputes; and on a range of other related issues. This is an example of the type of intervention that is best evaluated using a cluster randomised design. It is also typical of the social interventions currently being proposed for reducing health inequalities.¹⁰ Data from our study show how sample size calculations are affected by the group randomisation design. The results should also be applicable in the design of other studies of interventions to improve child and maternal health where general practices are to be randomised.

Statistical background

In a conventional randomised controlled trial it is routinely assumed that the data from each subject are independent of each other. Independence of data is a prerequisite for simple tests of significance such as *t* tests, and also for more complex statistical methods, such as regression. When clusters are randomised, the individuals within each group or cluster are unlikely to be independent of each other.

There are many possible reasons for this. If the cluster is a clinic or general practice, it might attract patients or clients who share similar attitudes or socioeconomic circumstances. If the intervention involves primary health care workers, their individual style and approach might profoundly affect the efficacy of the intervention. If the cluster is a school, its health ethos might influence the outcome measures of interest—for example, whether it has an antismoking policy. If the cluster is a neighbourhood, environmental factors are likely to affect all people in the neighbourhood similarly. The consequence of all these potential influences is that results from individuals within clusters tend to be more similar to each other than results from individuals between different clusters. The technical description of this phenomenon is intraclass correlation, and it can be measured by the intraclass correlation coefficient, or ρ .

Intraclass correlation reduces statistical power, which means that overall numbers of subjects have to be increased. Factors that need to be taken into account in sample size calculations are the distribution of the outcome measure in the population and the presumed effect of the intervention,¹⁰ the size of the individual clusters, the number of clusters, and the amount of intraclass correlation. It is intuitively obvious that increasing the number of clusters will increase the power, and that the lower the intraclass correlation, the greater the power. It is perhaps less obvious that simply increasing the number of subjects in each cluster will do relatively little to change the power because the increased precision of the results might be offset by a stronger effect of the intraclass correlation.

The ratio of the number required in a cluster randomised study to the number that would be required in an equivalent individually ran-

domised study is called the inflation factor. To calculate the inflation factor, the intraclass correlation coefficient of the outcome measure is needed.^{3 9 11 12} Several investigators have called for the publication of such data to inform the design of other studies in the same field.^{4 6}

Methods

SOURCE OF DATA

Six general practices from an urban area covering a mixed population took part. Three practices were randomly selected to provide Citizen's Advice Bureau services from within primary care, with the other three practices acting as controls. Eligible families were those with an infant under 1 year of age at the start of the trial and all new births during the nine months of the intervention. Two hundred and ninety families were recruited from these practices. We have included data on the mother and the first infant recruited in these families. We have omitted data from five families with more than one eligible infant because these data would complicate the intraclass correlation calculations. Outcomes included a range of measures of both maternal and child health derived from two questionnaires at the beginning and end of the intervention, and from examination of the general practice notes. From the questionnaires we selected for detailed consideration the Edinburgh postnatal depression score,¹³ questions about behaviour problems, sleeping problems, maternal smoking, and visits to the general practitioner in the previous two weeks, or to the hospital in the previous two months. The second questionnaire also included adapted questions on general health, common illnesses, behaviour, accidents, and hospitalisation from the Warwick child health and morbidity profile.¹⁴ From the general practice notes we extracted information on consultations, coded according to type of problem along lines developed by McKinney *et al.*¹⁵ Theoretically, data derived from the second questionnaire and the general practice records might show a spuriously high intraclass correlation because they are subject to the effect of the intervention; however, in practice, this will have had a minimal effect for reasons explained in the discussion.

The trial was conducted as a pilot for a projected larger definitive trial of the effects on maternal and child health of the provision of advice services. The objectives of this pilot were to provide data to calculate sample sizes, to identify the practical and logistic difficulties posed by such a trial, and to measure in detail the financial and material benefits obtained by mothers of young children who have access to a Citizen's Advice Bureau service within their doctors' surgeries.

STATISTICAL METHODS

Intraclass correlation coefficients for each of the outcome measures of interest were calculated according to the formula derived by Fleiss,¹⁶ which uses mean square values from a one way analysis of variance:

$$\rho = \frac{MS_{\text{Between}} - MS_{\text{Within}}}{MS_{\text{Between}} + (m-1)MS_{\text{Within}}} \quad (1)$$

where MS values are the mean squares from the ANOVA table and m is the average size of the clusters. This method is equally applicable to both dichotomous variables and to continuous variables.^{12–16} Because the size of each cluster varied, we substituted a term m_0 instead of m in the above equation¹⁷:

$$m_0 = \frac{1}{(k-1)} \left[n - \left(\frac{\sum m_i^2}{n} \right) \right] \quad (2)$$

where k is the total number of clusters, m_i is the number of cases in the “ i th” cluster, with i ranging from 1 to k , and n is the total number of individuals in the sample.

The inflation factor (IF) is calculated using the formula given by Donner *et al.*⁹

$$IF = 1 + (m-1)\rho \quad (3)$$

To calculate sample sizes, the number required in an individually randomised study is multiplied by the inflation factor. Sample sizes

for individually randomised studies were calculated using conventional formulae.¹⁸

Results

Table 1 shows data from the families who were eligible for our trial. Although the families of 919 children were eligible, only 290 families were recruited, 261 of whom returned the initial questionnaire, and 219 the second questionnaire. General practitioner notes were reviewed for 196 of the mothers and 200 of the children from the recruited families, the rest were unobtainable because they had moved or the notes were otherwise unavailable at the practice.

Table 2 shows the overall values of the health measures we collected for the total sample, the intraclass correlation coefficients calculated using equation 1 above, and a suggested minimum effect size, which would correspond to a clinically important outcome of a community based social intervention.

Table 3 shows the effect of increasing the number of subjects in each cluster on the total number of clusters required. As an example, we include an intervention to reduce the population mean Edinburgh postnatal depression score from 8.5 to 7.5. A conventional individu-

Table 1 Numbers of subjects with data from each practice and some socioeconomic data

	Practice					
	1	2	3	4	5	6
Eligible children	130	255	112	165	100	157
Recruited families	58	95	31	28	40	38
Questionnaire one returned	55	90	28	23	32	33
Questionnaire two returned	47	79	27	20	21	25
Data from GP notes on mother	44	65	21	18	20	28
Data from GP notes on child	44	67	21	18	22	28
Single parent families	12/57 (21%)	12/91 (13%)	5/30 (17%)	3/25 (12%)	8/33 (24%)	11/34 (32%)
Families on benefit	22/56 (39%)	23/91 (25%)	18/30 (60%)	7/24 (29%)	15/30 (50%)	15/34 (44%)
Home owners	24/46 (52%)	56/79 (71%)	10/25 (40%)	12/20 (60%)	6/21 (29%)	14/24 (58%)
No access to car	15/57 (26%)	14/91 (15%)	11/30 (37%)	2/25 (8%)	15/33 (45%)	10/34 (29%)
Overcrowded (> 1 person per room)	11/55 (20%)	19/89 (21%)	7/28 (25%)	6/23 (26%)	4/32 (13%)	9/33 (27%)

GP, general practitioner.

Table 2 Outcome measures, intraclass correlations, and suggested minimum effect sizes

Outcome measure	Mean (SD) and quartile values or prevalence (%)	Intraclass correlation coefficient (ρ)	Suggested size of minimum effect
EPDS (Q1)	8.4 (5.3) 4, 7, 12	0.029	8.5 → 7.5
EPDS ≥ 13 (Q1)	60/243 (24.7%)	0.0336	25% → 15%
EPDS (Q2)	7.4 (4.9) 4, 7, 11	0.0239	7.5 → 6.5
EPDS ≥ 13 (Q2)	31/211 (14.7%)	0.0108	15% → 10%
Mother smokes (Q1)	72/255 (28.2%)	0.0267	28% → 23%
Mother smokes (Q2)	53/213 (24.9%)	0.088	25% → 20%
Breast feeding if child under 6 months age	48/131 (36.6%)	0.0355	35% → 45%
Breast fed at birth	188/258 (72.9%)	0.0137	75% → 85%
Child has sleeping problem (Q1)	31/260 (11.9%)	0.0 (-0.0049)	12% → 8%
Child has sleeping problem (Q2)	27/218 (12.4%)	0.0 (-0.003)	12% → 8%
Child's general health—very healthy* (WCHMP)	106/216 (49.1%)	0.0287	50% → 60%
Minor illness—more than other children (WCHMP)	17/218 (7.8%)	0.0438	8.0% → 6.5%
> 2 minor illnesses in past 3 months (WCHMP)	41/213 (19.2%)	0.0479	19% → 15%
Accident in past year (WCHMP)	19/219 (8.7%)	0.0 (-0.0135)	9.0% → 7.5%
Behaviour problem (WCHMP)	14/216 (6.5%)	0.0206	6.5% → 4.5%
Hospitalised in past 6 months (WCHMP)	20/219 (9.1%)	0.0 (-0.0126)	9.0% → 7.5%
Child attended hospital in last two months	38/260 (14.6%)	0.0140	15% → 12%
Child seen GP in past 2 weeks (Q1)	87/259 (33.6%)	0.0 (-0.0196)	33% → 25%
Child seen GP in past 2 weeks (Q2)	50/219 (22.8%)	0.0 (-0.0161)	23% → 18%
Non-routine visits to GP by mother standardised to 1 year	3.2 (3.0) 0.9, 2.4, 4.6	0.0277	3.2 → 2.5
Non-routine visits to GP for child standardised to 1 year	4.2 (3.5) 1.6, 3.5, 5.9	0.0274	4.2 → 3.5

EPDS ≥ 13 indicates depression.

In cases where the calculated intraclass correlation coefficient is negative, by convention it is assumed to be equivalent to 0.

Questions that are duplicated in the second questionnaire are included to give an idea of the stability of our results.

*This is a positive measure of health, very healthy *v* healthy, not very healthy, or unhealthy.

EPDS, Edinburgh postnatal depression score; GP, general practitioner; Q, questionnaire; WCHMP, Warwick child health and morbidity profile.¹³

Table 3 Sample size requirements for a cluster randomised trial to reduce the average Edinburgh postnatal depression score from 8.5 to 7.5

Number in practice (m)	Inflation factor ($1 + (m - 1)\rho$)	Total sample size required	Number of practices
10	1.26	1112	112
20	1.55	1367	68
30	1.84	1623	54
40	2.13	1879	47
50	2.42	2134	43
75	3.15	2778	37
100	3.87	3413	34
150	5.32	4692	32
200	6.77	5971	30

This assumes 80% power to detect a difference with a two tailed significance of 95%. The number required in a conventional randomised controlled trial is 441 patients/arm (882 in total).

Figure 1 The effect of different cluster sizes on the number of subjects and number of clusters required for a cluster randomised trial. Solid line, number of subjects required; broken line, number of clusters required. The data illustrate a trial to reduce the average Edinburgh postnatal depression score from a mean value of 8.5 (SD, 5.3) to a mean of 7.5, with 80% power to detect the difference at a two tailed significance of 95%, assuming an intracluster correlation coefficient of 0.029.

ally randomised trial with 80% power to detect this difference with a two tailed significance of 95% would require 882 subjects (441 in each arm) using the population data from the first questionnaire in our study.¹⁸ If, instead, we conducted a trial where practices were randomised to provide the intervention, the table shows the total number of practices required when different sized samples from each practice are chosen. As the size in each cluster increases, the overall number of clusters reduces. However, there are diminishing re-

turns, in that there comes a point at which very large increases in cluster size are required to reduce the total number of clusters further. This is best seen graphically in fig 1.

We found that it would be possible to recruit 50 to 80 families with infants over a one year period from an average urban practice (table 1). Table 4 uses data from table 2 to show the number of practices that would be required for a cluster randomised trial to test the effects of a primary care based intervention on the outcome measures we have illustrated if 50 cases were recruited for each practice.

Discussion

We have shown that for a range of health outcomes in mothers and infants, the number of practices needed for a cluster randomised trial is large, and for some of these outcomes the number is too large to make such a trial feasible. The reasons that such large numbers are required are twofold. In some cases—for example, maternal smoking, breast feeding, and postnatal depression, there are relatively high intracluster correlation coefficients. This translates into a large inflation factor. It is not surprising that the intracluster correlations are high, because a range of social, cultural, and neighbourhood factors influence the characteristics of a practice population, and the approach of general practitioners, health visitors, and other primary care staff can potentially influence the health outcome measures.

In other cases—for example, sleeping problems and accidents, the intracluster correlations are low or non-existent. Here, large sample sizes are required because the underlying rate in the population is low, or the postulated effect of an intervention is small. This is not a problem for cluster based designs only: the same sample size difficulty would occur with an individually randomised design.

There are several drawbacks to our study. The small number of practices reduces the precision of our estimates of intracluster correlations. This has a potentially important impact. Small differ-

Table 4 Number of practices required for cluster randomised trial (CRT) of interventions for typical mother and child health outcomes, assuming 50 cases/practice

Measure	Effect size	Number for individual RCT	Inflation factor (IF)	Number of practices for CRT
EPDS (Q1)	8.5 → 7.5	882	2.42	43
EPDS ≥ 13 (Q1)	25% → 15%	540	2.65	29
EPDS (Q2)	7.5 → 6.5	754	2.17	33
EPDS ≥ 13 (Q2)	15% → 10%	1450	1.53	44
Mother smokes (Q1)	28% → 23%	2462	2.31	114
Mother smokes (Q2)	25% → 20%	2266	5.31	241
Breast feeding under 6 months	35% → 45%	790	2.74	43
Breast feeding at birth	75% → 85%	540	1.67	18
Sleep problem (Q1)	12% → 8%	1862	1	38
Sleep problem (Q2)	12% → 8%	1862	1	38
General health (WCHMP)	50% → 60%	814	2.41	39
Minor illness (WCHMP)	8.0% → 6.5%	9646	3.15	607
>2 minor illness (WCHMP)	19% → 15%	2866	3.35	192
Accidents (WCHMP)	9% → 7.5%	10824	1	216
Behaviour (WCHMP)	6.5% → 4.5%	3924	2.01	158
Hospitalised in past 6 months (WCHMP)	9% → 7.5%	10824	1	216
Hospitalised in past 2 months (Q1)	15% → 12%	4204	1.69	142
Seen GP in past 2 weeks (Q1)	33% → 25%	1058	1	21
Seen GP in past 2 weeks (Q2)	23% → 18%	2124	1	42
GP visits over 1 year (mother)	3.2 → 2.5	576	2.36	27
GP visits over 1 year (child)	4.2 → 3.5	785	2.34	37

These values assume 80% power to detect the stated difference with a two tailed significance of 95%.

EPDS, Edinburgh postnatal depression score; GP, general practitioner; Q, questionnaire; RCT, randomised controlled trial; WCHMP, Warwick child health and morbidity profile.

ences in the value of the intracluster correlation coefficient can translate into large differences in the calculated inflation factor. Second, the results apply to urban general practices covering a mixed population. It would be inappropriate to apply our results to sample size calculations that included rural practices or practices covering highly affluent or highly deprived populations. Another potential weakness is that we included both control and intervention practices in our study. If the intervention had a large effect, this would increase the apparent intracluster correlation. However, the intervention was poorly taken up and did not have a large demonstrable effect, and comparing intracluster correlation coefficients from both questionnaires shows no systematic increase over time (table 2). Despite the limitations of our results, they enable an informed approach to sample size calculations for cluster randomised trials in maternal and child health.

What then are the implications of our results? Cluster randomised trials based in primary care are feasible for some of the outcomes we have measured. Even where the inflation factor is two to three times for cluster sizes of 50, trial sizes are manageable. A good example would be a study of the provision of advice services to reduce postnatal depression. The real challenge is not to find outcomes with low intracluster correlation, but to find outcomes with a sufficiently high prevalence, and interventions with sufficiently large effect sizes. For example, sleeping and behavioural problems had a low prevalence in our study population, most of whom were under 1 year of age. If we had studied an older group of toddlers we might have expected a prevalence of disorders of around 30%. A trial of an intervention to reduce this prevalence to 20%, even assuming an inflation factor of 2.5, would not require more than 40 practices. The very low intracluster correlation we found in the younger age group might not be applicable to older children, but our results suggest that inflation factors are unlikely to exceed 2 to 3.

The publication of intracluster correlations within general practice populations of common child and maternal health problems is essential. Our results need to be seen in that context. We have presented the data in a way that enables different baseline assumptions about effect sizes and cluster sizes to be substituted in sample size calculations.

The more general implications of our results relate to the initial choice of study design. Where there is a choice between a cluster randomised design and a conventional individually randomised trial, the superficial attraction of having to deal with only a few clusters rather than many individuals might be outweighed by the increased overall size of the trial. From our experience on this pilot, there are many hidden difficulties in managing such studies. Where there is no choice, and a cluster randomised design is the only practical way of evaluating an intervention that maintains the benefits of randomisation, our results provide justification for the large size that might be necessary.

Appendix 1

The formula for ρ , the intracluster correlation coefficient, given in the text is the most convenient one for data that is available at the level of individuals. The value of ρ can also be calculated from data only available as rates or prevalences at the level of the cluster (that is, by general practice). Assume there are k clusters, x_i is the number of positive cases, and m_i the number of all cases in the "ith" cluster, with i ranging from 1 to k .

$$\rho = 1 - \frac{\sum \frac{x_i(m_i - x_i)}{m_i}}{k(\bar{m} - 1)(p(1 - p))}$$

In this equation, p is the overall proportion of positive cases and $\bar{m}$ is the mean cluster size. The results obtained in our study differ slightly if this equation is used, because the degrees of freedom for the mean square between clusters in this equation is taken as k rather than $(k - 1)$ in our study. We considered $(k - 1)$ to be the appropriate value because k was rather small in our study, whereas Fleiss assumes $k > 20$ for his derivation.¹⁶

The Citizen's Advice Bureau and family health study team are R Reading (paediatrician), M Mclean (health psychologist), J Appleby (health economist), S Reynolds (clinical psychologist), all from the school of health policy and practice, University of East Anglia; S Kember and J Watts, health visitor team leaders from the Norwich Community Health Partnership NHS Trust; and M Wright (assistant director) and J Wheatley (adviser) from the Norwich Citizen's Advice Bureau. The study was funded by the NHS Anglia and Oxford regional research and development directorate under grant number HSR/UEA/1196/87. We are grateful to all the families and health visitors, general practitioners, and practice managers for their involvement in the study. We also thank K Branson, M Busuttill, B Coupland, S Jarvis, S Muggford, T Peters, N Shih, S Steele, and J Tanner for advice, support, and data collection. The study received ethical approval from the Norwich local research ethics committee. Copies of the questionnaires are available from the authors on request.

- Oakley A, Roberts H, eds. *Evaluating social interventions*. Ilford: Barnardo's, 1996.
- Alderson P, Brill S, Chalmers I, et al. *What works? Effective social interventions in child welfare*. Ilford: Barnardo's, 1996.
- Koepsell TA, Wagner EH, Cheadle AC, et al. Selected methodological issues in evaluating community based health promotion and disease prevention programs. *Annu Rev Public Health* 1992;13:31-57.
- Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomisation, 1979-1989. *Int J Epidemiol* 1990;19:795-800.
- Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. *BMJ* 1998;317:1171-2.
- Slymen DJ, Hovell MF. Cluster versus individual randomisation in adolescent tobacco and alcohol studies: illustrations for design decisions. *Int J Epidemiol* 1997;26:765-71.
- Edwards SJL, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. *BMJ* 1999;318:1407-9.
- Underwood M, Barnett A, Hajioff S. Cluster randomisation: a trap for the unwary. *Br J Gen Pract* 1998;48:1089-90.
- Donner A, Birkett N, Buck C. Randomisation by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981;114:906-14.
- Department of Health. *Our healthier nation: a contract for health*. Cmnd 3854. London: The Stationery Office, 1998.
- Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *BMJ* 1998;316:1455.
- Murray DM, Rooney BL, Hannan PJ, et al. Intraclass correlation among common measures of adolescent smoking: estimates, correlates and application in smoking prevention studies. *Am J Epidemiol* 1994;140:1038-50.
- Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry* 1987;150:782-6.
- Spencer NJ, Coe C. The development and validation of a measure of parent-reported child health and morbidity: the Warwick child health and morbidity profile. *Child Care Health Dev* 1996;22:367-79.
- McKinney PA, Alexander FE, Nicholson C, Cartwright RA, Carrette J. Mothers' reports of childhood vaccinations and infections and their concordance with general practitioner records. *J Public Health Med* 1991;13:13-22.
- Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley and Son, 1981.
- Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford: Blackwell, 1994.
- Florey C duV. Sample size for beginners. *BMJ* 1993;306:1181-4.